



UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE CIÊNCIAS APLICADAS E EDUCAÇÃO
DEPARTAMENTO DE CIÊNCIAS EXATAS
BACHARELADO EM SISTEMAS DE INFORMAÇÃO

**DESCOBERTA DO CONHECIMENTO: A UTILIZAÇÃO DE DATA
MINING PARA ANÁLISE DE RECLAMAÇÕES DO SITE
CONSUMIDOR.ORG NO ESTADO DA PARAÍBA**

FRANCISCO VALDEVINO FERNANDES FAVARO
Orientador: Prof.^a Dr.^a Adriana Zenaide Clericuzi

Rio Tinto – PB
2016

FRANCISCO VALDEVINO FERNANDES FAVARO

**DESCOBERTA DO CONHECIMENTO: A UTILIZAÇÃO DE DATA
MINING PARA ANÁLISE DE RECLAMAÇÕES DO SITE
CONSUMIDOR.ORG NO ESTADO DA PARAÍBA**

Monografia apresentada para obtenção do título de Bacharel à banca examinadora no curso de Bacharelado em Sistemas de Informação no Centro de Ciências Aplicadas e Educação (CCAEE) – Campus IV da Universidade Federal da Paraíba.

Orientador(a): Prof.^a Dr.^a Adriana Zenaide Clericuzi

Rio Tinto – PB
2016

F272d Favaro, Francisco Valdevino Fernandes.

Descoberta do conhecimento: a utilização de Data Mining para análise de reclamações do site consumidor.org no Estado da Paraíba. / Francisco Valdevino Fernandes Favaro. – Rio Tinto: [s.n.], 2016.

57 f. : il.-

*Orientador (a): Profa. Dra. Adriana Zenaide Clericuzi.
Monografia (Graduação) – UFPB/CCAE.*

1. Algoritmo de associação. 2. Direito do consumidor. 3. Sistemas de informação.

*UFPB/BS-CCAE
004:347.4(043.2)*

CDU:

FRANCISCO VALDEVINO FERNANDES FAVARO

**DESCOBERTA DO CONHECIMENTO: A UTILIZAÇÃO DE DATA
MINING PARA ANÁLISE DE RECLAMAÇÕES DO SITE
CONSUMIDOR.ORG NO ESTADO DA PARAÍBA**

Trabalho de Conclusão de Curso submetido ao Curso de Bacharelado em Sistemas de Informação da Universidade Federal da Paraíba, Campus IV, como parte dos requisitos necessários para obtenção do grau de BACHAREL EM SISTEMAS DE INFORMAÇÃO.

Assinatura do Autor: _____

APROVADO POR:

Prof.^a Dr.^a Adriana Zenaide Clericuzi
Universidade Federal da Paraíba - CCAE - DCX

Prof. Dr. Joelson Nogueira de Carvalho
Universidade Federal da Paraíba - CCAE - DCX

Prof. Me. Rafael Marrocos Magalhães
Universidade Federal da Paraíba - CCAE - DCX

Rio Tinto – PB
2016

Dedico este trabalho a meu tio, Damião Fernandes Ferreira (In Memoriam), que em vida foi o exemplo de pai, me dando valores que levarei enquanto meu coração bater.

AGRADECIMENTOS

Agradeço primeiramente a Deus, pelo dom da vida e pela sede de conhecimento.

Agradeço a minha família, especialmente minha mãe Maria do Desterro Fernandes Ferreira e minha avó Josefa Fernandes Ferreira, pilares em minha vida e em quem me sustento.

Agradeço a minha namorada Ana Paula Queiroz, pelo zelo e cuidado durante tantas noites em claro e momentos de completa tristeza, onde suas palavras foram motivacionais para que esse momento fosse possível.

Agradeço a minha querida e amiga Prof.^a Adriana Zenaide Clericuzi, que com paciência e seriedade dedicou-se a me auxiliar na produção deste trabalho, dando-me exemplo de ensino, cuidado e generosidade, valores que levarei pelo resto de minha vida.

Agradeço a querida e amiga Professora Maika Zampier, minha primeira mentora na universidade e figura fundamental na minha formação pessoal, profissional e acadêmica.

Agradeço ao Mestre Paulo Roberto Palhano, homem de índole, caráter e sabedoria, de quem absorvi conhecimentos e conselhos e a quem agradeço pelo crescimento na área de pesquisa e no desejo da profissão de ensino.

Agradeço aos amigos que inicialmente me acolheram em seu meio, quando não tinha ninguém com quem contar: Alexandro Sátiro, Hugo Campos e Riackson Menezes. Pessoas essenciais que me apoiaram nos momentos iniciais da graduação.

Agradeço aos amigos Ermeson Nóbrega e Nathane Henrique, que suportaram meus estresses de perto durante toda a produção deste trabalho. Pessoas especiais que levarei pelo resto da minha vida. E aos amigos Lais Soares, Rivaldo Júnior, Sérgio Euclides, Adriano Martins, André Luís, Ana Paula Pontos, Tatiana Barbosa, Lumma Rabay, Severino Ramos, Samuel, Nil Alisson, Anderson Alves, Victor Demétrio, pessoas que estarão em meu coração enquanto ele bater, sempre serão lembrados pelos momentos de alegria que foram compartilhados durante a academia.

Agradeço aos professores do Departamento de Ciências Exatas e de outros departamentos, com os quais tive o prazer de conviver e pelos quais recebi o conhecimento que hoje levarei para o exterior dos muros da universidade.

Por fim, mas não menos importante, gostaria de agradecer aos amigos e/ou conhecidos que, direta ou indiretamente, tiveram participação na produção deste trabalho.

RESUMO

Vivemos em um mundo onde as informações são produzidas mais rápido do que podem ser analisadas. O conjunto de informações pode ser considerado um tesouro, se bem usado pela organização, mas normalmente, essas informações estão sendo escritas e não utilizadas. Neste contexto, surge o *Knowledge Discovery in Databases* ou descoberta do conhecimento em bases de dados, que realiza a análise de informações em bancos de dados com o intuito de descobrir conhecimentos ocultos. O KDD conta com uma série de etapas e ferramentas, que suportam todo o processo, desde a escolha do banco ao momento do descobrimento do conhecimento e interpretação dos resultados. Relacionando os bancos de dados das organizações, podemos observar que um grande volume de dados é produzido pelo governo, mas boa parte de suas análises não tornam-se públicas, sendo essas de total interesse da população. Sendo assim, esse trabalho objetiva aplicar a técnica de KDD, focando especialmente no processo de Data Mining, com o intuito de descobrir possíveis regras de associação em um banco de dados de reclamações do site Consumidor.gov.br na Paraíba, buscando avaliar sua eficiência como ferramenta de prestação de serviço e auxílio no processo de resolução de problemas.

Palavras-chave: KDD, Data Mining, Descoberta do conhecimento, Regras de Associação.

ABSTRACT

We live in a world where information is produced faster than they can be analyzed. The set of information can be considered a treasure, although used by the organization, but usually this information is being written and unused. In this context, the Knowledge Discovery in Databases or knowledge discovery in databases arises, which performs the analysis of information in databases in order to discover hidden knowledge. KDD has a series of steps and tools that support the entire process, from choosing the bank at the time of the discovery of knowledge and interpretation of results. Linking the databases of organizations, we can see that a large volume of data is produced by the government, but much of their analyzes do not become public, and these total interest of the population. Thus, this work aims at applying the KDD technique, focusing especially on data mining process, in order to discover possible association rules in a database of Consumidor.gov.br site complaints in Paraíba, seeking to evaluate their efficiency as service and support delivery tool in the problem-solving process.

Keywords: KDD, Data Mining, Discovery Knowledge, Association Rules.

LISTA DE FIGURAS

Figura 01: Ambiente Inicial do site ReclameAqui.....	21
Figura 02: Hierarquia de Dados, Informação e Conhecimento.....	27
Figura 03: Etapas do KDD.....	28
Figura 04: Tarefas da Mineração de Dados.....	30
Figura 05: Interface Inicial do Software Weka.....	32
Figura 06: Exemplo de Arquivo ARFF.....	33
Figura 07: Arquivo de Dados CVS Reclamações.....	34
Figura 08: Arquivo CSV no LibreOffice Calc	35
Figura 09: Tabela de Dados Organizada	36
Figura 10: Arquivo ARFF resultante do processo de Transformação	37
Figura 11: Dados carregados no WEKA	39
Figura 12: Resultado da Aplicação do Algoritmo de Associação	39

LISTA DE GRÁFICOS

Gráfico 01: Relação Número de Regras x Nível de Confiança.....	40
Gráfico 02: Relação Reclamações Finalizadas e Avaliadas x Reclamações Resolvidas.....	40
Gráfico 03: Procurou Empresa x Finalizada e Avaliada x Respondida	41
Gráfico 04: Relação Sexo Masculino x Finalizada e Avaliada x Respondida	41
Gráfico 05: Relação entre Área de Telecomunicações e Respondidas	42
Gráfico 06: Relação João Pessoa x Finalizada e Avaliada x Respondida.	43
Gráfico 07: Relação Cobrança/Contestação x Respondidos	43
Gráfico 08: Relação Sexo Masculino x Procurou Empresa x Finaliza e Avaliada x Respondida.....	44
Gráfico 09: Relação Telecomunicações x Procurou Empresa x Respondido.....	44
Gráfico 10: Relação entre João Pessoa x Procuraram Empresa x Finalizada e Avaliada x Respondida	45
Gráfico 11: Relação Cobrança/Contestação x Procurar Empresa x Respondido	45

LISTA DE QUADROS

Quadro 01: Resultados dos Objetivos Específicos do Trabalho	47
Quadro 02: Resultados das Questões de Pesquisa do Trabalho	47

LISTA DE SIGLAS

API – Application Programming Interface

ARFF – Attribute-Relation File Format

BI – Business Intelligence

CDC – Código de Defesa do Consumidor.gov.br

CSV – Comma-separated values

GUI – Graphical User Interface

JVM – Java Virtual Machine

KDD – Knowledge Discovery in Databases

ODM – Oracle Data Mining

P&G – Proctor and Gamble

SAS – Statistical Analysis System

SPSS – Statistical Package for the Social Sciences

UF – Unidade Federativa

WEKA – Waikato Environment for Knowledge Analysis

SUMÁRIO

1. INTRODUÇÃO	1
1.1. Considerações Iniciais:	1
1.2. Justificativa:	2
1.3. Objetivo Geral:	2
1.4. Objetivo Específico:	3
1.5. Questões de Pesquisa:	3
1.6. Metodologia:	3
2. FUNDAMENTAÇÃO TEÓRICA	5
2.1. O Consumidor e sua relação de aquisição	5
2.2. O código de Defesa do Consumidor	6
2.3. Vício em Produto e Defeito	7
2.4. A proteção ao consumidor e sua Relação direta com a Internet	7
2.5. Dados Abertos: Consumidor.gov.br	10
3. ASPECTOS ESSENCIAIS: A MINERAÇÃO DE DADOS E SEUS CONCEITOS.	12
3.2. Business Intelligence (BI)	12
3.2. Dado, Informação e Conhecimento: O início da Tomada de Decisão.	13
3.3. A Descoberta do Conhecimento: KDD (Knowledge discovery in databases)	14
3.4. Data Mining	16
3.5. Ferramentas de Mineração de Dados	18
4. A APLICAÇÃO DA MINERAÇÃO DE DADOS	20
4.2. O Software Weka	20
4.2. Execução das Etapas da Descoberta do Conhecimento: Seleção e Processamento	21
4.3. Transformação	25
5. DATA MINING: A UTILIZAÇÃO DE ALGORITMO DE ASSOCIAÇÃO	27
5.2. A aplicação do Algoritmo de Associação	27
6. RESULTADOS	36
7. TRABALHOS FUTUROS	38
8. CONCLUSÃO	39
REFERÊNCIAS	40

1. INTRODUÇÃO

Neste capítulo será abordada a temática do trabalho, fazendo uma explanação inicial sobre o processo de Tomada de Decisão Gerencial e como o *Business Inteligente* (B.I.) pode auxiliar o gerente neste processo, realizando a aplicação de técnicas específicas de *Data Mining* para obtenção de conhecimento.

1.1. Considerações Iniciais:

Vivemos na era da Informação, e essa informação é produzida de forma diária e constante. A informação é parte essencial de nosso dia e muitas vezes a construímos de forma empírica, sem o conhecimento de que aquela ação é na verdade a construção de uma informação e que essa informação pode ser transformada em conhecimento e posteriormente utilizada em estratégias de negócio.

Pequenos e médios negócios e empreendimentos lidam com a informação como algo trivial, e a organizam em seus bancos de dados, mas essa organização não utiliza dados estatísticos que são gerados por essas informações como possíveis geradores de qualidade e determinantes de vendas de produtos. Armazenar esses dados e não utilizá-los de forma correta ou não utilizá-los e pode ser considerada uma má estratégia da gerência, que deve ter nesses dados as bases para o processo de tomada de decisão dentro de organizações.

Uma das principais fontes de conhecimento e informação é o governo, esfera de poder que regulamenta diversos segmentos de mercado nacionais e tem por objetivo proporcionar o bem estar aos cidadãos, transformando seus impostos em bens e serviços públicos. Mas o governo, além disso, ainda tem como função a transparência de suas informações, tornando possível que a população faça uma completa análise dos dados que são gerados por seus diversos órgãos, mas a análise e aprofundamento nesse grande volume de dados pode gerar uma série de problemas, quando aos métodos utilizados e se esses métodos e resultados gerados serão de confiança.

Visando resolver este problema, diversos métodos e técnicas, estatísticas e empíricas, foram desenvolvidas com o objetivo de auxiliar e dar-lhe suporte com dados que possam revelar informações que anteriormente estavam invisíveis ou simplesmente não eram aproveitadas corretamente. Essas técnicas se utilizam de algoritmos de Inteligência Artificial

e dados estatísticos e geram relatórios estratégicos, gráficos para referenciarão e/ou indicadores preditivos e utilizam com base o banco de dados das organizações.

Com o crescente número de bancos de dados virtuais e a constante necessidade de técnicas de análise e classificação da informação para apoio a obtenção de novo conhecimento e novas informações, este trabalho apresentará uma das mais utilizadas técnicas nas grandes organizações para o apoio no processo de tomada de decisão, o *Data Mining* (Mineração de Dados), visando realizar uma análise da eficiência do site consumidor.gov como solucionador ou apoiador da sociedade na solução de suas reclamações.

1.2. Justificativa:

Com a evolução da sociedade, tem emergido uma vasta gama de ferramentas de apoio ao processo de descoberta do conhecimento e transformação desse conhecimento em informação. Essas ferramentas fornecem a sociedade um conjunto de informações sobre o ambiente dos dados que serão estudados, bem como diversas formas de análise e um grande número de resultados, para cada técnica aplicada.

Muitas organizações a exemplo de *Walmart*, *Proctor and Gamble* (P&G) tem realizado investimentos em técnicas de análise de seus bancos de dados, com o objetivo de descobrir padrões e prever tendências futuras para o mercado em seus respectivos segmentos. Sendo assim, torna-se visível que a obtenção de conhecimento é essencial para a vitalidade e lucratividade de organizações e para a melhoria de segmentos quando nos referimos ao governo ou informações que são relacionadas diretamente com o bem-estar da sociedade.

Tendo em vista o cenário apresentado e a importância da análise dos bancos de dados das organizações para obtenção de sólido conhecimento, apoiado pelos Sistemas de Informação, este trabalho se propõe a apresentar a técnica de *Data Mining* como ferramenta de análise para descoberta de padrões afim de realizar uma análise da eficiência do site consumidor.org como ferramenta de resolução de problemas, tendo como base para isso os dados de reclamações fundamentadas do primeiro trimestre de 2016 no estado da Paraíba.

1.3. Objetivo Geral:

Aplicar uma técnica de *Data Mining* no conjunto de informações sobre reclamações de consumidores durante o primeiro trimestre de 2016 no estado da Paraíba e apresentar a

relevância da técnica para avaliação do ambiente na resolução de problemas e auxílio ao consumidor.

1.4. Objetivo Específico:

Para que seja possível alcançar com sucesso o objetivo deste trabalho, é necessário realizar uma série de procedimentos que são essenciais no processo de constituição dos resultados. São eles:

- Realizar Análise do Banco e Determinar a técnica de *Data Mining* mais adequada, utilizando como ferramenta o Software Weka;
- Realizar o processo de Mineração de Dados, seguindo as etapas prescritas e organizadas no KDD (Knowledge Discovery in Databases), descrevendo os resultados com a aplicação;
- Apresentar os Resultados da Aplicação da Técnica de Associação;
- Observar a solução dos problemas com base nas regras de associação e demonstrar a eficiência do ambiente em apoiar o consumidor;

1.5. Questões de Pesquisa:

- QP01. Qual a real importância do *BI* para obtenção de novos conhecimentos?
- QP02. Os consumidores podem considerar o ambiente do consumidor.gov como um forte meio de solução de seus problemas e comunicação com as organizações reclamadas?
- QP03. Podemos avaliar a eficiência de órgãos com base em seus dados públicos ?

1.6. Metodologia:

Com o intuito de satisfazer com plenitude a proposta apresentada neste trabalho, pode-se descrever o processo de concepção desta pesquisa nas seguintes etapas:

Etapa I: Revisão Bibliográfica: Análise de livros, artigos, periódicos, e outros materiais impressos ou digitais com temática relacionada.

Etapa II: Revisão histórica: Como as técnicas de *Data Mining* e tomada de decisão foram utilizadas em organizações no decorrer da história.

Etapa III: Obtenção do Banco de Dados para análise;

Etapa IV: Análise e Organização dos dados;

Etapa V: Seleção do Algoritmo de mineração de dados adequado para o banco selecionado;

Etapa VI: Aplicação da Técnica de Mineração dos Dados;

Etapa VII: Identificação dos padrões e Apresentação dos resultados;

Etapa VIII: Apresentar os dados e correlacioná-los com a eficiência do órgão em atender as demandas da sociedade.

Como ferramenta, este trabalho utilizará o Software de Business Intelligence WEKA, uma aplicação desenvolvida utilizando a tecnologia Java e contendo um vasto pacote de algoritmos de Inteligência Artificial. Os diferenciais para a escolha dessa ferramenta foram: Ser gratuita e possuir código aberto.

2. FUNDAMENTAÇÃO TEÓRICA

2.1. O Consumidor e sua relação de aquisição

Historicamente, nos deparamos com relações de compra e venda desde o início dos registros históricos. É visível que sempre houve compra e venda de produtos ou serviços, caracterizada pela troca de produtos por metais preciosos ou pela troca de um produto por outro.

Com o avanço da história, novos conceitos de aquisição de produtos surgiu-se e o perfil dos consumidores transformou-se com o avanço da história e dos produtos. Essa mudança observou-se ser mais representativa a partir do século XVIII com o acontecimento da Revolução Industrial, transformando completamente o modelo de indústria conhecido pela sociedade. Após a revolução, surgiu um intermediário entre o processo de compra e venda, e o conceito de aquisição tornou-se universalizado, ou seja, todos precisavam de tudo.

Após a revolução tecnológica, surgiram novos e revolucionários paradigmas industriais, a sociedade então deparou-se com um valoroso número de produtos e serviços em circulação no mercado, graças a ideia de produção em massa. Rocha (1992) ressaltou que graças a esses acontecimentos, a segurança dos consumidores diminuiu consideravelmente em decorrência de erros técnicos e falhas nos processos de produção em massa.

O objetivo geral da sociedade industrial foi gerar lucro para si, mas durante esse processo observou-se que a qualidade dos produtos e serviços oferecidos não condizia com valores cobrados, afetando assim a relação existente entre os fornecedores e os consumidores.

Vendo a séria problemática, observou-se a necessidade por parte dos fornecedores do desenvolvimento de métodos de passar ao consumidor uma falsa sensação de qualidade que seria adquirida junto com o produto ou serviços que o mesmo compraria.

Com o avanço desses problemas, e as sérias consequências que estavam ocorrendo aos consumidores, pensou-se na necessidade de um conjunto de regras que tivesse por objetivo proteger o consumidor no momento da aquisição de produtos e/ou serviços que apresentassem problemas, garantindo seus direitos, e determinando uma série de deveres que consumidor e fornecedor deveriam seguir com o objetivo de garantir a melhor qualidade ao consumidor. Nesse processo, surge o Código de Defesa do Consumidor.

2.2. O código de Defesa do Consumidor

O código de defesa do consumidor (CDC), promulgado pela lei nº 8.078/1990, mas estudiosos afirmam que seu surgimento deu-se de forma tardio no sistema legislativo brasileiro. Nunes (2009) afirma que a primeira lei de defesa do consumidor nos Estados Unidos surgiu em 1890, e que até o surgimento da lei brasileira, os consumidores deveriam amparar-se no Código Civil de 1916, que era caracterizado como individualista e patrimonialista.

Durante o período pré CDC era um processo comum que as indústrias explorassem a mão de obra, se negasse de forma contratual a realizar o pagamento de possíveis indenizações e inviabilizassem a concorrência, garantindo o monopólio de seus produtos. Muitas ainda praticavam o Cartel, realizando a combinação de valores dos produtos, garantindo que ela e seus concorrentes conseguissem obter certo nível de lucro.

Mas pode ser vista uma vantagem nesse processo de atraso na concepção de um conjunto de normas específicas para proteção ao consumidor, já que durante o desenvolvimento do projeto de lei que criou o CDC, foram inseridas leis inspiradas em países que já possuíam um sólido conjunto de regulamentações, evitando, em parte, a necessidade de criação de novos paradigmas legais. Por sua solidez, o CDC já foi inspiração para desenvolvimento de políticas direcionadas ao consumidor na Argentina, Uruguai, Paraguai e países da União Europeia.

O CDC assim, surge como ferramenta máxima de proteção e amparo de consumidores, desde o processo de contratação e/ou aquisição de produtos ou serviços até o momento de sua utilização, podendo o mesmo analisar sua valia e a possibilidade do mesmo ser considerado não ideal, seja por não ter a qualidade esperada, seja por seu preço demasiado alto para um produto considerado inferior.

Nery (1999) considera a relação de consumo como sendo “a relação jurídica existente entre fornecedor e consumidor tendo como objeto a aquisição de produtos ou utilização de serviços pelo consumidor”. Podemos assim analisar a relação de consumo como um instrumento jurídico e legal, embasado por leis e protegido diretamente pelo CDC.

Com a implantação do CDC na década de 90, fortes mudanças entre a relação consumidor e fornecedor surgiram, e novas métricas no processo de produção foram criados,

visando gerar mais qualidade de produto e mais segurança na sua utilização pelo consumidor.

Nunes (2009) relata:

“Dentro dessa estratégia geral dos negócios, como fruto da teoria do risco, um item específico é o que está intimamente ligado à sistemática normativa adotada pelo CDC. É aquele voltado à avaliação da qualidade do produto e do serviço, especialmente a adequação, finalidade, proteção à saúde, segurança e durabilidade. Tudo referendado e complementado pela informação”. (NUNES, 2009).

Vemos assim o CDC com não apenas um avanço para a legislatura, mas para todo um contexto social, melhorando o processo de relação entre cliente e fornecedor e apresentando critérios que qualifiquem o produto ou serviço para comercialização.

2.3. Vício em Produto e Defeito

O CDC determina produto como sendo “qualquer bem material ou imaterial, móvel ou imóvel”. Ainda podemos observar a classificação de produtos como duráveis ou não duráveis, sendo considerados produtos duráveis aqueles que têm sua vida útil prolongada para favorecer sua constante utilização, não perdendo sua eficiência com o consumo. Por sua vez, os produtos não duráveis, são aqueles que perdem eficiência no momento que sua utilização se encerra.

Já sabendo a classificação de produto, podemos então determinar que vício o conjunto de características, sejam de qualidade ou quantidade que tornem o produto ou serviço inadequado para consumo ao qual são destinados.

São constantes e estão cada dia mais elevados, o número de reclamações e a utilização do CDC com o objetivo de coibir práticas de vício em produtos. Muitas vezes, essas práticas são maquiadas, ou a qualidade do produto torna-se inferior por defeito no processo de produção.

Já o defeito é considerado um acontecimento que causa um dano superior ao dano do vício, é algo extra, considerado assim como um acontecimento devastador e acarretando prejuízo financeiro e efeito catastrófico no produto e/ou serviço.

2.4. A proteção ao consumidor e sua Relação direta com a Internet

Com a evolução da sociedade e em paralelo a evolução da tecnologia, a internet passou a ser não apenas um meio de propagação de informação, mas um meio de criação de informação e comunicação. Consumidores com acesso à internet tem se tornado ávidos

pesquisadores no tocante ao processo de aquisição de produtos e serviços, tanto de forma presencial, como em lojas virtuais.

Como criadora de informação, atualmente é possível realizar a análise e verificação da qualidade de muitas empresas brasileiras e estrangeiras por meio de uma simples pesquisa em ferramentas de busca, tais como Google e Yahoo. Com isso, surgiu uma geração de consumidores precavidos, pesquisadores de qualidade e avaliadores de conteúdo. Surgiu a geração dos Consumidores Digitais.

Mas, cabe ao texto uma observação: Faz-se necessário a visibilidade de consumidores que não tem meios de acesso a informação na internet, mas que buscam informações de qualidade com outros consumidores, são pessoas que se comunicam entre si com o intuito de obter a melhor informação sobre o produto mais bem qualificado de acordo com os que usaram e/ou contrataram o produto e/ou serviço.

É possível encontrar uma vasta quantidade de fóruns, blogs, vídeos, imagens e textos, todos realizando avaliação de empresas, de produtos adquiridos, lojas onde compraram serviços que contrataram, etc. Essas avaliações são base para que muitos usuários realizem suas operações de compra e/ou contratação.

Mas, tendo em vista que o comércio eletrônico de produtos cresceu de forma exponencial no século XX e início do XXI, viu-se a necessidade de fazer com que o Consumidor Digital também se transformasse em um Reclamante Digital, mas que essa reclamação pudesse ser realizada em ambiente propício para isso, contando com o apoio de pessoas que entendessem do problema, e que essa reclamação pudesse ser direcionada diretamente a empresa/pessoa responsável pelo produto/serviço do qual se está reclamando. Neste contexto, surgiram os ambientes de reclamação virtuais.

O mais famoso dos ambientes virtuais de reclamação é o site ReclameAqui (<http://www.reclameaqui.com.br/>). Nele é possível visualizar a classificação das empresas cadastradas, a avaliação dos usuários, um ranking com as empresas que mais resolvem problemas, um ranking com as piores empresas dos últimos 30 dias ou do ano inteiro. Enfim, uma infinidade de possibilidades que fazem com que o consumidor possa pesquisar a reputação da empresa onde está adquirindo algo, com base nas informações e reclamações de um conjunto de usuários de todo o Brasil.

O site ReclameAqui foi idealizado em 2001, depois de seu criador ter problemas com uma empresa aérea e não encontrar um canal apropriado para a formalização de sua

reclamação e a difusão dessa reclamação para outras pessoas. Atualmente o ambiente conta com mais de 15 milhões de usuários cadastrados e com mais de 120 mil empresas cadastradas e tem um total de mais de 600 mil acessos diários.

Abaixo, é possível observar o ambiente inicial do site ReclameAqui:

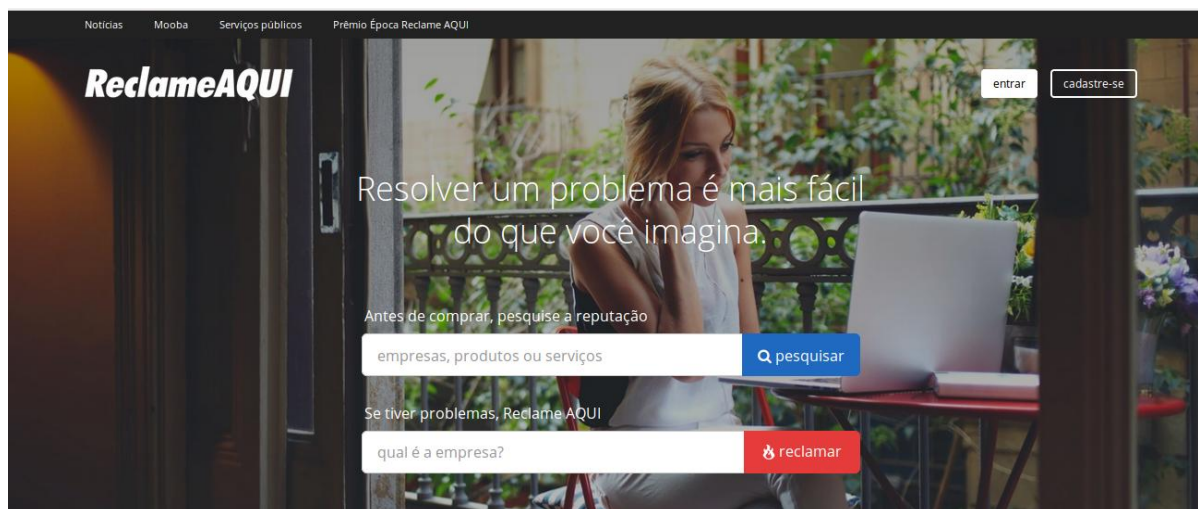


Figura 01: Ambiente Inicial do site Reclame Aqui.

Mais recentemente, o governo brasileiro, com o intuito de gerar dados mais específicos sobre reclamações em âmbito local e que são registradas via internet, lançou o ambiente chamado Consumidor.gov.br. Esse ambiente conta com as mesmas ferramentas e funcionalidades do reclame aqui, mas com um diferencial de que as empresas podem escolher participar deste ambiente ou não.

As empresas não são obrigadas a aderirem, mas caso o façam, assinam um contrato assegurando o comprometimento de visualizar, analisar, responder e solucionar toda e qualquer reclamação que seja registrada neste ambiente, levando sempre em consideração o contexto do problema, e as possibilidades da organização.

Com as informações que são registradas no site, o governo pode gerar dados para buscar melhoria no conjunto de políticas para proteção ao consumidor, incentivo a competitividade de mercado e melhoria na qualidade e no atendimento ao consumidor.

As informações depositadas no site são compartilhadas pelo portal de dados abertos do governo e monitoradas diretamente pelos Procons e pela Secretaria Nacional do Consumidor, que está ligada ao Ministério da Justiça, e tem apoio direto da sociedade.

São observadas 3 políticas nas quais o ambiente se baseia para buscar a melhor comunicação possível entre cliente, fornecedor e estado, sendo elas:

1. Transparência e controle social são imprescindíveis à efetividade dos direitos dos consumidores;
2. As informações apresentadas pelos cidadãos consumidores são estratégicas para gestão e execução de políticas públicas de defesa do consumidor;
3. O acesso a informação potencializa o poder de escolha dos consumidores e contribui para o aprimoramento das relações de consumo.

Levando em consideração essas políticas, podemos observar a construção de um ambiente que visa o total compartilhamento de suas Informações com a sociedade e que visa utilizar essas informações para benefício do povo. O cidadão, então, pode realizar a investigação sobre lojas antes de realizar a aquisição de qualquer produto e/ou serviço.

Ambos os ambientes de reclamação necessitam de um cadastro, com informações pessoais, e uma descrição clara e direta do problema que está sendo relatado. A reclamação é direcionada para a empresa, mas a mesma pode ser observada por qualquer pessoa na rede, esteja ela logada no ambiente de reclamação ou não, ou seja, a informação de reclamação é pública.

2.5. Dados Abertos: Consumidor.gov.br

Por apresentar um conjunto sólido e confiável de informações, este trabalho utilizou-se de um conjunto de dados do Consumidor.gov.br, referente as reclamações impetradas na plataforma no período correspondente de Janeiro a Março de 2016. Essas informações estão disponíveis no ambiente de dados abertos do governo e podem ser acessadas por qualquer pessoa com acesso à internet.

A escolha da base de dados deu-se, não apenas pelos fatores citados acima, mas por sua relevância, sendo um conjunto real e atual de reclamações, de diversos segmentos e sobre diversos fatores. Ainda foi considerado como fator predominante na escolha da base de dados, a facilidade para a obtenção da informação, já que a mesma é pública e aberta.

Como característica do conjunto de dados, podemos observar o número de critérios, sendo ao total 21 e o conjunto de informações, totalizando mais de 52 mil linhas. Realizada a filtragem das informações, neste trabalho utilizaremos 419 linhas, que são o conjunto de

informações oficiais e fundamentadas na plataforma por usuários paraibanos no período acima descrito.

O conjunto de informações está organizado no formato Comma-separated values (CSV), que proporciona a troca de informações entre uma tabela e aplicativos, facilitando assim a manipulação da mesma para análise de suas informações.

3. ASPECTOS ESSENCIAIS: A MINERAÇÃO DE DADOS E SEUS CONCEITOS.

3.2. Business Intelligence (BI)

Historicamente falando, povos antigos como fenícios e persas utilizavam sua inteligência para correlacionar informações que tinham como fonte a natureza, com o objetivo de determinar se ia chover, se a maré estava propícia a navegação, ou outros fatores determinantes para a época. Assim, desde a antiguidade o homem tem realiza o processo de análise das informações, buscando descobrir padrões ou determinar situações futuras de acordo com o conjunto de dados disponível. Com essas informações os povos antigos buscavam melhorias para seus grupos e comunidades.

Sezões, Oliveira e Baptista (2006) determinam que o BI tem como meta “recolher dados, transformá-los em informação (através de descoberta de padrões e tendências) e, sequencialmente, informação em conhecimento útil e oportuno para a tomada de decisão.” Podemos assim determinar que o BI tem ligação direta com os processos de coleta, armazenamento e análise dos dados de uma organização, facilitando o processo de tomada de decisão e auxiliando gerentes e interessados a obter conhecimento útil.

Barbieri(2001) destaca que BI pode ser considerada um termo “guarda-chuva” onde são determinadas ferramentas e técnicas que auxiliam no processo de tomada de decisão, sendo considerado um processo que visa obter a informação certa, no momento mais oportuno possível.

Com a evolução das tendências no mercado, as informações a cada dia precisam ser analisadas e diretamente cruzadas. Vendo essa necessidade, é impossível desassociar BI tomada de Decisão e Ferramentas de Aprendizado de Máquina. Segundo Petrini *et al* (2005) BI pode ser determinada como:

“A assertiva segundo a qual, somente quando se têm 80% do quebra-cabeça é que se pode começar a ver coisas que as outras pessoas não veem. Juntando, fazendo encaixes, reorganizando informações e conhecimentos fragmentados, dispersos e aparentemente sem sentido, pode-se chegar a conclusões inéditas, mesmo que estas informações já estivessem disponíveis, porém não disponibilizadas de maneira tal que fosse possível enxergar algo mais” (PETRINI *et al*, 2005).

Sendo assim, é possível observar que o BI não é uma ferramenta, mas sim um conceito que pode ser diretamente vivenciado no ambiente organizacional, sendo ele um diferencial para o processo competitivo entre organizações. O conceito de BI nas

organizações, então, enquadra-se pelo contexto de que é impossível de forma manual classificar e organizar a grande quantidade de informação que diariamente é produzida e há a necessidade de utilização de ferramentas de Tecnologia da Informação que possam, de forma efetiva e eficiente, realizar a extração de conhecimento desse vasto conjunto de informações.

Com a aplicação dos conceitos de BI, uma organização pode vislumbrar conhecimentos antes não visíveis, e aplicá-los na organização como diferencial em relação a seus competidores diretos, mas ainda é visível a falha nas organizações que adotam BI, focando na metodologia de avaliação de desempenho e não à geração e flexibilização da informação.

Batista (2004) apresenta BI: “As ferramentas de BI podem fornecer uma visão sistêmica do negócio e ajudar na distribuição uniforme dos dados entre os usuários, sendo seu objetivo principal transformar grandes quantidades de dados em informações de qualidade para a tomada de decisões. Através delas, é possível cruzar dados, visualizar informações em várias dimensões e analisar os principais indicadores de desempenho empresarial”.

Observamos assim que as ferramentas de Business Intelligence, tais como Weka, Pentaho Community e outras são consideradas de BI por utilizarem-se de um grupo de dados e transformá-los em informação/conhecimento. Os dados são retornados em forma de gráficos, curvas e/ou tabelas apresentando relações e/ou resultados das análises comparativas.

Com o panorama e as considerações observadas, podemos então observar que a utilização de BI nas organizações é um ativo que agrega valor ao negócio, diminuindo custos e aumentando receitas, definindo padrões, tendências e dando subsídios para a criação e/ou modificação de produtos.

3.2. Dado, Informação e Conhecimento: O início da Tomada de Decisão.

As organizações atualmente produzem mais dados do que conseguem analisar. Esses dados, em sua maioria organizados em um banco digital, são considerados de difícil interpretação por parte dos Analistas de Informação. Sendo assim, há a necessidade de processos que subsidiem a obtenção de conhecimento nessas bases e dados e gere informação válida para que o gerente possa tomar a decisão, com base nas opções que lhe são apresentadas.

Rezende (2003) e Mota (2004) apresentam a ideia de interligação e entre dado, informação e conhecimento, sendo os três um conjunto indissociável. O dado é algo bruto que pode ser tratado e qualificado, a informação é um parâmetro que tem utilização no

processo de análise dos dados, e por fim, o conhecimento como a capacidade de geração de modelos com o intuito de solucionar um problema na organização, apresentando o cenário do problema e propondo a(s) possíveis solução(ões).

Rezende (2003) ainda considera o dado qualquer tipo de mídia ou informação que possa ser processada computacionalmente. Sendo assim analisamos a informação como um conjunto de dados que são, classificados e interpretados, observando o surgimento de possíveis métricas ou padrões e associações entre os mesmos. Decisão, sendo assim, pode ser considerada como a atividade de utilização de um determinado conhecimento adquirido anteriormente em processo de análise de dados e extração de informação.

Na figura 02 abaixo podemos vislumbrar a hierarquia de dado, informação e conhecimento, inseridos em uma pirâmide. Interpretando a imagem, vemos os dados com a base, a informação bruta e sensível, que depois de analisada torna-se informação. A informação por sua vez é estudada e sobre ela são gerados padrões e métricas, chamados de conhecimento. Esse conhecimento resultante do processo de análise dos dados é o que será utilizado na tomada de decisão.



Figura 02: Hierarquia de Dados, Informação e Conhecimento (Alesso, 2004)

3.3. A Descoberta do Conhecimento: KDD (Knowledge discovery in databases)

O termo KDD (*Knowledge Discovery in Databases*) foi formalizado em 1989 e faz referência a constante busca de informações em grande volume de dados. KDD é um conjunto de processos que ocorre em várias etapas, tendo como objetivo a identificação de padrões

compreensíveis, válidos, novos e potencialmente úteis a partir de grandes conjuntos de dados (Fayyad, 1996). O processo é considerado iterativo por necessitar de repetição nas etapas de KDD em determinados momentos do processo, e é considerado não trivial por apresentar um considerável nível de complexidade de execução.

Para que a aplicação do KDD em uma base de dados aconteça de forma correta é necessário que o analista que executará a atividade no banco de dados tenha sólidos conhecimentos da diferença entre dado, informação e conhecimento – estes no âmbito de tecnologias de informação. Goldshmidt & Passos (2005) classificam dado como itens elementares e sem relação que são captados e armazenados por algum recurso de tecnologia da informação.

Com a alta complexidade envolvida no processo de KDD em bases de dados, há a necessidade de análise e planejamento de atividades. Essa análise e planejamento deve definir “o que fazer” com relação a base de dados, definindo objetivos e seguindo o conjunto de atividades apresentada por Boente, Goldschmidt & Estrela (2006), sendo essas atividades: Identificar pessoas e áreas envolvidas com a aplicação, Realizar o levantamento sobre o software e hardware existente, inventar bases de dados disponíveis, analisar bases de dados e compreender o significado de relevância dos atributos avaliando a qualidade e quantidade dos dados disponíveis, desenvolver uma lista de potenciais expectativas por parte das pessoas e áreas, e identificar e documentar o conhecimento prévio.

O processo de KDD divide-se em diversas fases, essenciais para o processo de descoberta do conhecimento, sendo elas: Seleção, pré-processamento, transformação ou formatação, Mineração dos Dados e avaliação. A etapa de seleção consiste no processo de apresentação do conjunto de dados que serão minerados, podendo esses dados estarem organizados ou não e provenientes do sistema de banco de dados da organização; a etapa de pré-processamento é o processo de tornar os dados adequados de acordo com a aplicação e a possível informação que se deseja obter; a etapa de transformação ou formatação consiste no processo de adequação dos dados de acordo com a aplicação, ou seja, os dados são organizados e tabelas e informações sem valor ou com possíveis incorreções e que não são essenciais são retirados; Após essas etapas ocorre a mineração dos dados, sendo eles analisados utilizando o algoritmo específico para a aplicação (esse processo será explicado

detalhadamente mais a frente); por fim, existe a apresentação dos resultados e avaliação, mostrando o conhecimento que foi adquirido de todo o processo de KDD.

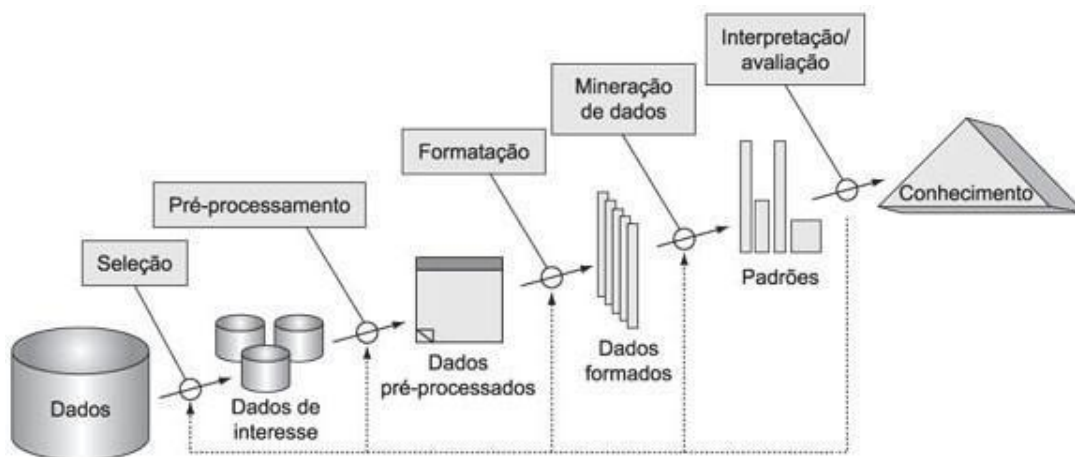


Figura 03: Etapas do KDD (Adaptado de Fayyad, 1996)

Na Figura 3 é possível visualizar graficamente o processo de KDD descrito anteriormente e suas etapas, desde os dados brutos e não tratados ao conhecimento, ferramenta que suporta a decisão.

3.4. Data Mining

Data Mining ou Mineração de Dados é a etapa do KDD responsável pela análise e descoberta de padrões e informações que são, em sua maioria, invisíveis sem a utilização de um algoritmo específico que transforme os dados em conhecimento utilizável. Esses dados, após sua extração e mineração, tornam-se um auxiliar para que o gerente tenha informações o suficiente e consequentemente o conhecimento de que decisão tomar para resolver o problema que está sendo analisado na Tomada de Decisão.

Cardoso (2008) conceitua Data Mining como sendo:

“uma técnica que faz parte de uma das etapas da descoberta de conhecimento em banco de dados. Ela é capaz de revelar, automaticamente, o conhecimento que está implícito em grandes quantidades de informações armazenadas nos bancos de dados de uma organização. Essa técnica pode fazer, entre outras, uma análise antecipada dos eventos, possibilitando prever tendências e

comportamentos futuros, permitindo aos gestores a tomada de decisões baseada em fatos e não em suposições.” (CARDOSO, 2008)

Fayyad (1996) conceitua mineração de dados da seguinte maneira: “Extração de conhecimento de base de dados é o processo de identificação de padrões válidos, novos, potencialmente úteis e compreensíveis embutidos nos dados”. Com o objetivo de compreender mais profundamente esta definição, Rezende (2003) realiza uma análise das principais palavras e as explica, com o objetivo de enriquecer e contextualizar Fayyad. Contextualizando cada palavra, temos:

1. Dados: Conjunto de informações presentes em um repositório ou banco de dados.
2. Padrões: Determinado subconjunto de dados apresentando pós utilização de método declarativo.
3. Processo: Conjunto de Etapas, tais como descoberta dos dados, análise de padrões e avaliação do conhecimento.
4. Válidos: Os padrões descobertos devem ser avaliados como aceitáveis, respeitando limites e objetivos definidos no processo de planejamento da aplicação.
5. Novos: os padrões descobertos devem possuir algum grau de certeza, devem satisfazer funções ou limiares que garantam que os exemplos cobertos e os casos relacionados ao padrão sejam aceitáveis.
6. Úteis: Os padrões, após descobertos, são incorporados para utilização futura.
7. Compreensíveis: Os padrões devem ser descritos em alguma linguagem, proporcionando compreensão por parte do usuário interessado.
8. Conhecimento: o conhecimento é definido em função do seu domínio de aplicação, utilidade, originalidade e compreensão.

A figura 4, apresentada abaixo, exemplifica a relação existente entre os possíveis resultados na pesquisa de Data Mining, que envolve a descoberta do conhecimento para um determinado fim. Resultados preditivos estão mais ligados a eventos futuros e posicionamentos, enquanto os descritivos expõem características, perfis e agrupamentos de dados.

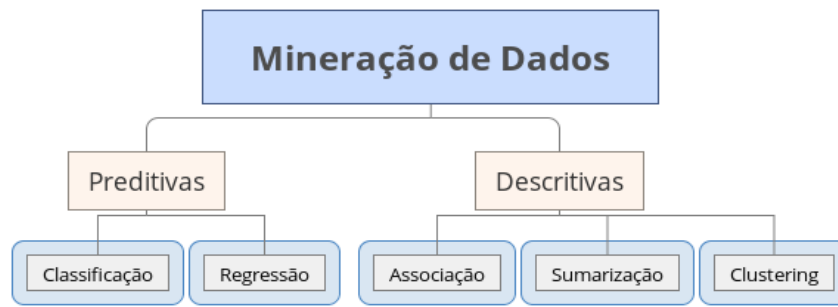


Figura 04: Tarefas da Mineração de Dados

São tarefas preditivas:

- Classificação: Classifica em classe dados anteriormente não classificados, criando para isso um modelo. Através da classificação é possível realizar previsões e antecipar tendências.
- Estimativa / Regressão: Define valor, para uma variável aparentemente desconhecida.

São tarefas Descritivas:

- Associação: Grupos que itens que ocorrem simultaneamente, gerando assim um número considerável de regras. (Lacerda & Souza, 2004)
- Clusterização: Divisão de determinado grupo heterogêneos em subgrupos mais homogêneos.
- Sumariação: tem por finalidade encontrar uma descrição simplificada a determinado subgrupo de dados. Sumarização, normalmente, é utilizada para criação de relatórios automáticos e análise exploratória (Fayyad, 1996).
- Dependência: Verifica se existe um modelo que descreva a dependência existente entre as várias variáveis analisadas.

A técnica de Data Mining utiliza-se de algoritmos de inteligência artificial e Aprendizado de Máquina, para extrair o máximo de informações úteis e válidas para apoiar o gerente ou diretoria no processo de tomada de decisão.

3.5. Ferramentas de Mineração de Dados

No mercado, atualmente, existe uma vasta diversidade de ferramentas com o intuito de realizar tarefas de mineração de dados e aprendizado de máquina. A escolha uma ou várias dessas ferramentas dependerá diretamente do tipo de resultado ou atividade que se pretende

desenvolver. Abaixo serão apresentadas as mais conhecidas ferramentas de mineração de dados disponíveis no mercado:

- Clementine: Uma das ferramentas líderes no mercado, desenvolvida pela Statistical Package for the Social Sciences (SPSS).
- SAS Enterprise Miner Suite: Desenvolvida pela Statistical Analysis System (SAS), essa ferramenta proporciona ao seu utilizador trabalhar em todas as etapas do processo de mineração utilizando módulos.
- WEKA: Ferramenta Open Source, desenvolvida pela Universidade de Waikato. Utilizaremos essa ferramenta na aplicação da Mineração apresentada em seguida.
- Oracle Data Mining (ODM): Desenvolvida pela Oracle e feita exclusivamente para seus bancos.
- IBM Intelligent Miner: Desenvolvida pela IBM e utilizada em Bancos DB2.

4. A APLICAÇÃO DA MINERAÇÃO DE DADOS

4.2. O Software Weka

O Software utilizado no desenvolvimento deste trabalho foi o WEKA, uma aplicação Open Source e gratuita, desenvolvida inicialmente em 1993 na universidade de Waikato, e sendo adquirido por uma empresa privada em 2006. Por possuir seu código fonte aberto, é possível realizar modificações em seus algoritmos, de acordo com a necessidade de seu usuário. A aplicação funciona em linha de comando e pode ser executada em sua interface gráfica padrão (GUI), favorecendo o aprendizado de usuários.

Na figura 05 abaixo, é possível observar a interface padrão do Software Weka:



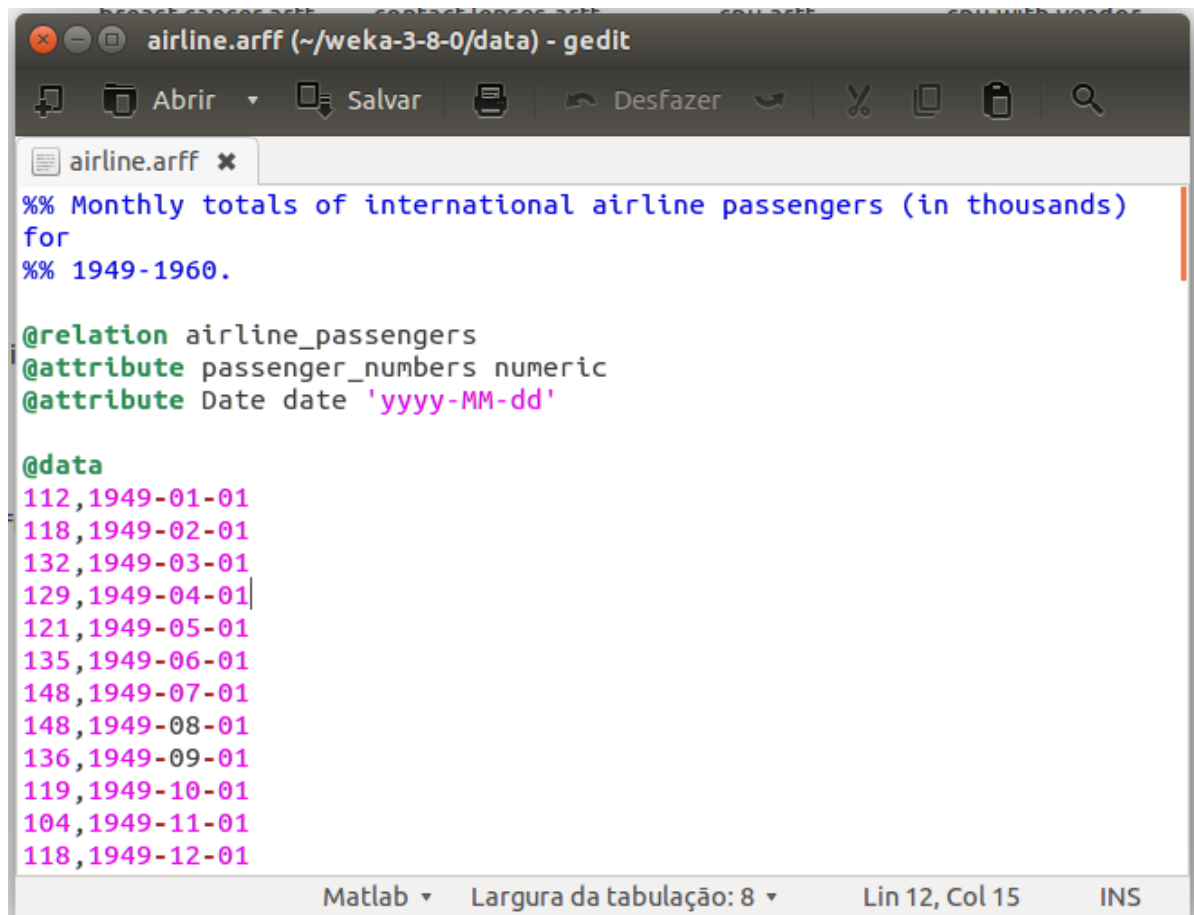
Figura 05: Interface Inicial do Software Weka

O Weka é escrito em linguagem Java, sendo portátil para qualquer Sistema Operacional que conte com uma Java Virtual Machine (JVM). O objetivo principal da aplicação é agregar um conjunto de algoritmos de Inteligência Artificial dedicado ao estudo da área de Aprendizado de Máquina. O Weka ainda conta com uma API completa, sendo possível incorporar os algoritmos em suas aplicações e realizar, por exemplo, atividades de mineração do lado do servidor, apenas enviando os dados pelo lado cliente.

Antes de aplicar os dados a qualquer algoritmo do pacote Weka estes devem ser convertidos para o formato ARFF que consiste basicamente de duas partes. A primeira contém uma lista de todos os atributos, onde devemos definir o tipo do atributo ou os valores

que ele pode representar, quando utilizamos os valores estes devem estar entre “{ }” separados por vírgulas. A segunda parte consiste das instâncias, ou seja, os registros a serem minerados com o valor dos atributos para cada instância separado por vírgula, a ausência de um item em um registro deve ser atribuída pelo símbolo “?”.

A figura 06 abaixo exemplifica um modelo estrutural do arquivo ARFF disponibilizado como teste no próprio WEKA.



```
airline.arff (~/.weka-3-8-0/data) - gedit
Abrir Salvar Desfazer
airline.arff x
%% Monthly totals of international airline passengers (in thousands)
for
%% 1949-1960.

@relation airline_passengers
@attribute passenger_numbers numeric
@attribute Date date 'yyyy-MM-dd'

@data
112,1949-01-01
118,1949-02-01
132,1949-03-01
129,1949-04-01
121,1949-05-01
135,1949-06-01
148,1949-07-01
148,1949-08-01
136,1949-09-01
119,1949-10-01
104,1949-11-01
118,1949-12-01
Matlab Largura da tabulação: 8 Lin 12, Col 15 INS
```

Figura 06: Exemplo de Arquivo ARFF

4.2. Execução das Etapas da Descoberta do Conhecimento: Seleção e Processamento

Como anteriormente denominado, o processo de seleção consiste na seleção do conjunto de dados que serão utilizados no processo de descoberta do conhecimento. Esses dados podem vir de diversas fontes e nos mais diversos formatos, sendo necessário em determinados momentos o desenvolvimento de uma aplicação específica para realizar a carga dos dados.

No processo de desenvolvimento deste trabalho, foram consultados o conjunto de dados aberto do governo federal. Nesse ambiente virtual (<http://dados.gov.br/>) são disponibilizados conjuntos de informações de todos os órgãos do governo e, em 3 anos, existe o planejamento de disponibilizar 100% de conteúdo de informações das esferas Federal, Estadual e Municipal.

Durante o processo de consulta, foi decidido utilizar como banco de dados o conjunto de reclamações que foram fundamentadas no sistema do ambiente virtual Consumidor.gov.br , um ambiente especializado na comunicação entre as empresas e os consumidores, visando proporcionar uma solução mais rápida dos problemas que o consumidor possa vir a ter com produtos ou serviços.

O arquivo de extensão .csv está dividido com as tabelas: Região, UF, Cidade, Sexo, Faixa Etária, Ano Abertura, Mês Abertura, Data Abertura, Data Resposta, Data Finalização, Tempo Resposta, Nome Fantasia, Segmento de Mercado, Área, Assunto, Grupo Problema, Problema, Como Comprou ou Contratou, Procurou Empresa, Respondida, Situação, Avaliação Reclamação, Nota do Consumidor, Total.

A figura 07 abaixo apresenta a parte inicial do arquivo CSV contendo as informações que serão processadas.

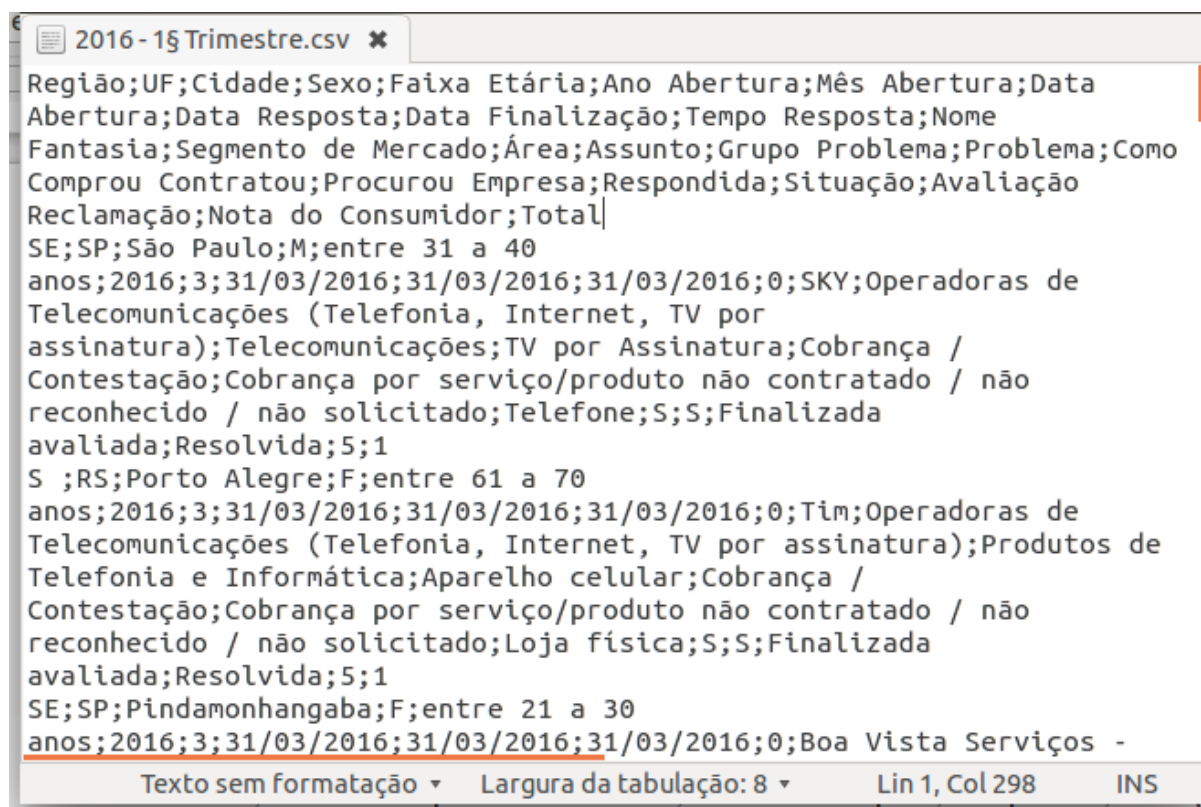


Figura 07: Arquivo de Dados CVS Reclamações.

O conjunto de informações foi filtrado e as informações que não tinham referência com o estado da Paraíba foram removidas. A figura 08 abaixo apresenta o conjunto de informações já no formato de tabela para aplicativos de planilhas eletrônicas, tais como Excel, da Microsoft ou Calc, do LibreOffice.

	A	B	C	D	E	F	G	H
1	Região	UF	Cidade	Sexo	Faixa	Etária	Ano	Abertura
2	NE	PB	Pilõesinhos	M	at		20 anos	2016
3	NE	PB	Arara	M	at		20 anos	2016
4	NE	PB	Bayeux	M	at		20 anos	2016
5	NE	PB	Cabedelo	M	at		20 anos	2016
6	NE	PB	Cajazeiras	M	at		20 anos	2016
7	NE	PB	Queimadas	F	at		20 anos	2016
8	NE	PB	São Sebastião de		Lagoa	de	Rocha	M
9	NE	PB	Alhandra	F	entre		41 a	50 anos
10	NE	PB	Alhandra	F	entre		31 a	40 anos
11	NE	PB	Alhandra	F	entre		31 a	40 anos
12	NE	PB	Areial	F	entre		41 a	50 anos

Figura 08: Arquivo CSV no LibreOffice Calc

Neste contexto, foi observado que há uma distorção na criação das tabelas, contendo dados desorganizados e informações em tabelas não condizentes com seu respectivo tipo. Podemos ver, por exemplo, na linha 8 que o nome da cidade transpassa sua coluna normal e consequentemente os demais dados estão em colunas que não são as corretas.

Após a Análise inicial e classificação correta dos dados na tabela, foi possível vislumbrar de forma mais clara o conjunto de dados. A Figura 09 abaixo apresenta um prospecto do conjunto de informações organizadas.

	A	B	C	D	E	F
1	Região	UF	Cidade	Sexo	Faixa Etária	Ano
2	NE	PB	Alagoa Nova	M	Até 20 anos	2016
3	NE	PB	Alagoa Nova	M	Até 20 anos	2016
4	NE	PB	Alagoa Nova	F	Até 20 anos	2016
5	NE	PB	Alagoa Nova	M	Entre 41 a 50	2016
6	NE	PB	Alagoa Nova	F	Até 20 anos	2016
7	NE	PB	Alagoa Nova	M	Até 20 anos	2016
8	NE	PB	Alagoa Nova	F	Até 20 anos	2016
9	NE	PB	Alagoa Nova	F	Entre 31 a 40	2016
10	NE	PB	Alagoa Nova	F	Até 20 anos	2016
11	NE	PB	Alhandra	M	Entre 51 a 60	2016
12	NE	PB	Alhandra	F	Entre 31 a 40	2016
13	NE	PB	Alhandra	M	Entre 41 a 50	2016
14	NE	PB	Arara	F	Entre 51 a 60	2016
15	NE	PB	Areial	M	Entre 21 a 30	2016
16	NE	PB	Bayeux	M	Entre 21 a 30	2015
17	NE	PB	Bayeux	M	Entre 21 a 30	2016
18	NE	PB	Bayeux	M	Entre 21 a 30	2016
19	NE	PB	Bayeux	M	Entre 21 a 30	2016
20	NE	PB	Bayeux	M	Entre 21 a 30	2016
21	NE	PB	Bayeux	M	Entre 21 a 30	2016
22	NE	PB	Bayeux	M	Entre 21 a 30	2016

Figura 09: Tabela de Dados Organizada.

4.3. Transformação

Após a seleção e organização dos dados existe o processo de transformação desses dados em um conjunto legível pela aplicação que utilizaremos no experimento. Esse processo consiste, então, na transformação das tabelas de informação para um Arquivo ARFF (Attribute-Relation File Format).

A estrutura do arquivo ARFF é composta por um conjunto de informações, sendo eles @Relation, @Attribute e @Data. O @Relation é o nome da relação de dados que utilizaremos. Nesta aplicação utilizaremos o nome de “Dados_Reclamações”. Os @Attributes são o as colunas da nossa tabela, ou tais como Região, Cidade, Etc. Os @Attributes são basicamente classificados em Nominiais ou Numéricos, mas ainda é possível, para algumas aplicações específicas, classificá-los como Data, para datas do calendário.

O arquivo desta aplicação contem um total de 21 @Attributes, entre numéricos e nominiais. Por não serem relevantes para a análise dos dados, as colunas da tabela “Região” e “Estado” não foram transformados em @Attributes no nosso arquivo ARFF, pois seria redundante já que os dados fazem menção a região e estado, e nesta aplicação estamos lidando apenas com dados locais.

Por fim, temos o campo @Data, que é composto pelo grupo de informações da tabela. As informações são preenchidas sequencialmente, de acordo com a ordem de declaração dos @Attributes e separados por vírgula. Ao total, o arquivo gerado com base na tabela de informações do processo de seleção formou um total de 419 linhas de @Data.

A figura 10 abaixo apresenta o resultado do processo de transformação dos dados e o arquivo ARFF resultante deste processo.

```

2016 - 1º Trimestre... Estado Paraiba.arff
%Francisco Favaro

@RELATION Dados_Reclamações

@ATTRIBUTE cidade
{Alagoinha,Alhandra,Arara,Areal,Bayeux,Boqueirão,Cabedelo,Cajazeiras,
@ATTRIBUTE sexo {F,M}
@ATTRIBUTE faixaetária {21-30,31-40,41-50,51-60,61-70,<20,>70}
@ATTRIBUTE anoabertura REAL
@ATTRIBUTE mêsabertura REAL
@ATTRIBUTE dataabertura
{01/01/2016,01/02/2016,01/03/2016,01/12/2015,02/01/2016,02/02/2016,02/03/
@ATTRIBUTE dataresposta
{01/02/2016,01/03/2016,02/01/2016,02/02/2016,02/03/2016,03/03/2016,04/01/
@ATTRIBUTE datafinalização
{01/01/2016,01/02/2016,01/03/2016,02/01/2016,02/02/2016,02/03/2016,03/02/
@ATTRIBUTE temporesposta REAL
@ATTRIBUTE nonefantasia
{AmericanasCom,AssurantSeguradora,AtivosSA,AviancaOceanair,AzulLinhaAérea
(TecBan),BancoBradesco,BancoDoBrasil,BancoItaúUnibanco,BancoPan
(Panamericano),BancoSantander,BancoSantanderCartões,BoaVistaServiçosSCPC
(ServiçoCentralDeProteçãoAoCrédito),Bradescard,BradescoCapitalização,Brad
Embratel,ClaroTV,Crediar,DecolarCom,Dotz,Electrolux-
LinhaBranca,ExtraCom,GVT,GollinhasAéreas,HSBC,HiperBompreço,Hipercard,Ita
Avent/Saeco/
Sonicare,Polishop,PontofrioCom,PositivoInformática,SKY,Samsung,Shoptime,S
Telefônica,WalmartCom}
@ATTRIBUTE segmentomercado
{AgênciasDeViagens,ArtigosEsportivos,Bancos/Financeiras/
AdministradorasDeCartão,BancosDeDados/

Matlab  Largura da tabulação: 8  Lin 1, Col 18  INS

```

Figura 10: Arquivo ARFF resultante do processo de Transformação

5. DATA MINING: A UTILIZAÇÃO DE ALGORITMO DE ASSOCIAÇÃO

Quando nos direcionamos para a análise e desenvolvimento de regras de associação em bancos de dados, podemos considerar o algoritmo *Apriori* um dos mais conhecidos e clássicos do processo de extração do conhecimento.

Concebido inicialmente pelo grupo de desenvolvimento do *Intelligent Miner* da IBM, esse algoritmo tem como finalidade resolver problemas relacionados com *Itemsets*.

O algoritmo divide-se na realização de duas atividades: Geração de *Itemsets* e Geração das regras de Associação. Tem por função a resolução de problemas referente a conjunto de informações que se relacionam entre si e que são frequentes no banco de dados.

Baseando-se nessas informações, podemos então deduzir que o algoritmo utiliza-se dos *Itemsets* já existentes e de suas relações para que novos *Itemsets* possam ser gerados, utilizando como base para essas informações o conhecimento prévio das propriedades dos *Itemsets* existentes.

Gerando uma relação lógica dentro da aplicação, podemos então definir que o algoritmo encontra o primeiro conjunto de *Itemsets*, e utiliza os encontrados para realizar a busca por novos *Itemsets* na aplicação. O Último conjunto de *Itemsets* encontrado sempre será utilizado como base para obtenção do próximo conjunto, até não existirem mais *Itemsets* a serem encontrados.

Ainda, existe a delimitação de um limite mínimo de confiabilidade que é determinada pelo usuário no momento da execução do algoritmo, sendo assim, a aplicação a ser analisada procurará *Itemsets* com confiabilidade igual ou superior a determinada pelo usuário.

5.2. A aplicação do Algoritmo de Associação

Após a realização da carga das informações no WEKA, é possível vislumbrar o conjunto de dados carregado para a aplicação do algoritmo. A imagem 11 abaixo é possível vislumbrar os dados organizados dentro do WEKA.

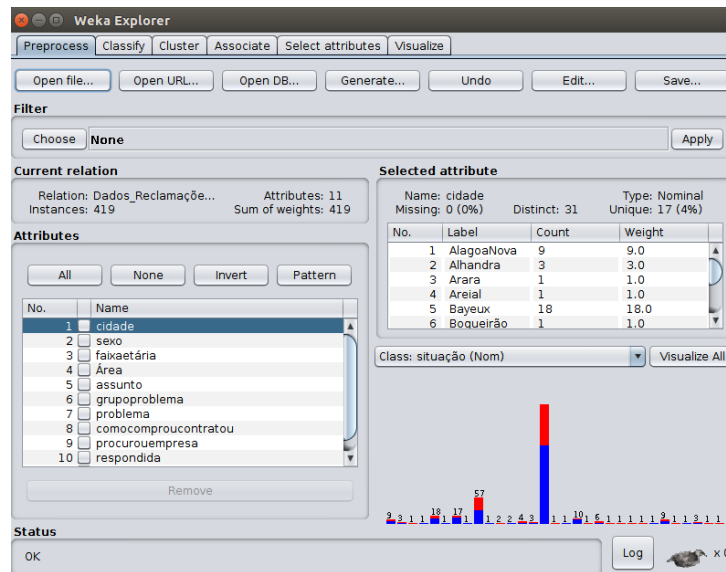


Figura 11: Dados carregados no WEKA

É possível visualizar o conjunto de informações de cada um dos atributos do arquivo, expressos pelo conjunto de colunas representadas na parte direita do aplicativo. O conjunto de informações que serão inseridas na base de dados irá ser apresentada na aba *Associate* onde deverá ser selecionado o algoritmo *Apriori* e onde o mesmo deverá apresentar as regras de associações.

Após a inserção dos dados, delimitação de confiança *Confidence* no valor de 0.9 e a solicitação da criação de 10 regras de associação do conjunto de dados, foi possível realizar a verificação do seguinte conjunto de informações. É possível visualizar o resultado inicial da aplicação do algoritmo na Figura 12 abaixo:

```

1. situação=FinalizadaAvaliada 264 ==> respondida=S 264 <conf:(1)> lift:(1) lev:(0) [1] conv:(1.26)
2. procurouempresa=S situação=FinalizadaAvaliada 237 ==> respondida=S 237 <conf:(1)> lift:(1) lev:(0) [1] conv:(1.13)
3. sexo=M situação=FinalizadaAvaliada 188 ==> respondida=S 188 <conf:(1)> lift:(1) lev:(0) [0] conv:(0.9)
4. Área=Telecomunicações 174 ==> respondida=S 174 <conf:(1)> lift:(1) lev:(0) [0] conv:(0.83)
5. cidade=JoãoPessoa situação=FinalizadaAvaliada 171 ==> respondida=S 171 <conf:(1)> lift:(1) lev:(0) [0] conv:(0.82)
6. grupoproblema=Cobrança/Contestação 170 ==> respondida=S 170 <conf:(1)> lift:(1) lev:(0) [0] conv:(0.81)
7. sexo=M procurouempresa=S situação=FinalizadaAvaliada 169 ==> respondida=S 169 <conf:(1)> lift:(1) lev:(0) [0] conv:(0.81)
8. Área=Telecomunicações procurouempresa=S 158 ==> respondida=S 158 <conf:(1)> lift:(1) lev:(0) [0] conv:(0.75)
9. cidade=JoãoPessoa procurouempresa=S situação=FinalizadaAvaliada 153 ==> respondida=S 153 <conf:(1)> lift:(1) lev:(0) [0] conv:(0.73)
10. grupoproblema=Cobrança/Contestação procurouempresa=S 152 ==> respondida=S 152 <conf:(1)> lift:(1) lev:(0) [0] conv:(0.73)

```

Figura 12: Resultado da Aplicação do Algoritmo de Associação

Durante a realização do Experimento foi possível observar que o número de regras de associação aumentava a medida em que o nível de confiança solicitado era decrescido. Abaixo é possível observar um gráfico que demonstra essa relação:

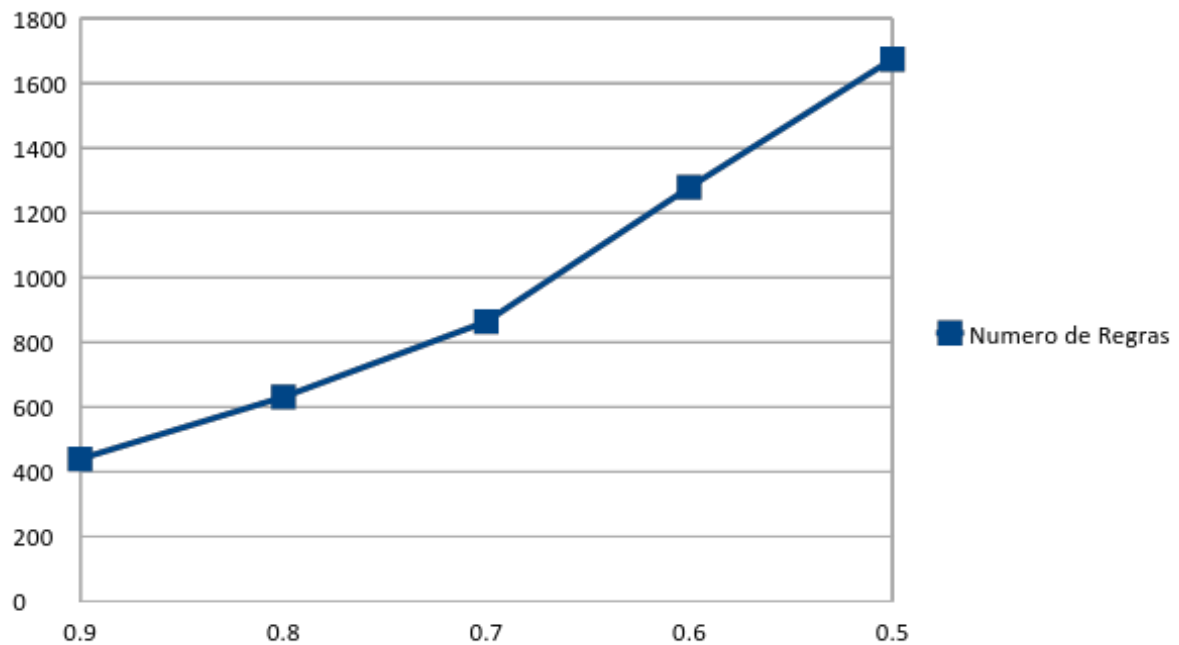


Gráfico 01: Relação Número de Regras x Nível de Confiança

Ao final do processo, podemos observar a geração de 15 regras de associação com um grau de confiança entre 99% e 100%. A maioria das regras de associação podemos observar foram geradas com reação a tabela do sexo.

Avaliando a primeira regra de Associação, podemos observar que 264 pessoas que Finalizaram e Realizaram a avaliação da reclamação tiveram sua reclamação resolvida. Essa relação pode ser expressa no gráfico abaixo:

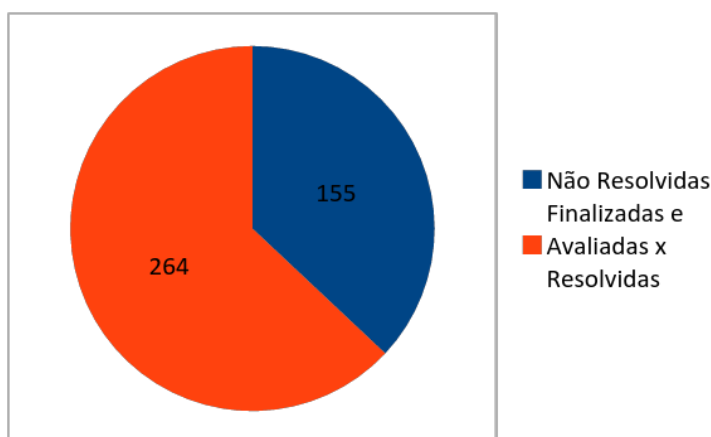


Gráfico 02: Relação Reclamações Finalizadas e Avaliadas x Reclamações Resolvidas

Avaliando a Segunda Regra de Associação podemos observar que o grupo representado pelas pessoas que procuraram a empresa antes de realizar a reclamação, que Finalizou e Avaliou a reclamação foi respondido pela empresa reclamada no momento da inserção da reclamação. Sendo assim, temos o resultado da relação entre Procurou Empresa x Finalizada e Avaliada x Respondida gera um total de 237 usuários. Essa avaliação pode ser observada no gráfico 3 abaixo:

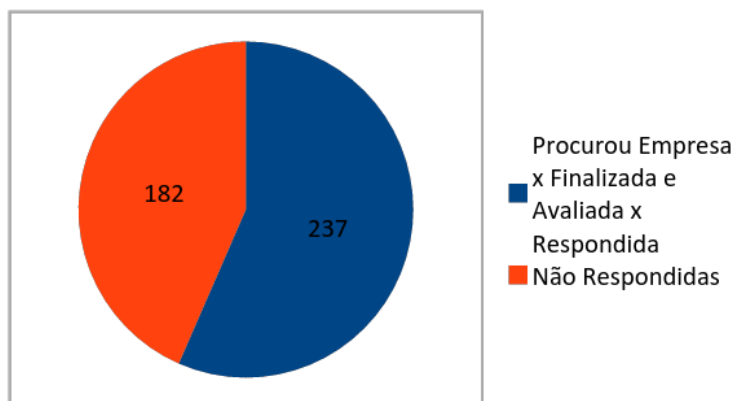


Gráfico 03: Procurou Empresa x Finalizada e Avaliada x Respondida

Podemos observar na terceira regra de associação a relação existente entre Sexo, Situação e se foi respondida. Na aplicação percebeu-se que o resultado dessa relação foi que 188 usuários que são do sexo Masculino e que tem como Situação: Finalizada e Avaliada, foram respondidos pelas empresas no momento em que a buscaram. Abaixo o gráfico 04 que representa a relação existente e o resultado em relação ao total de usuários:

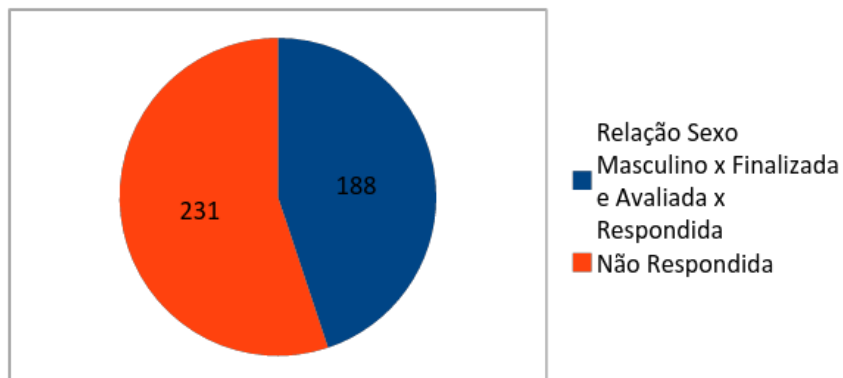


Gráfico 04: Relação Sexo Masculino x Finalizada e Avaliada x Respondida

Na quarta regra de associação podemos observar a relação entre a área da empresa reclamada, neste caso, Telecomunicação e a informação que foi respondida. Sendo assim, o algoritmo apresenta um total de 174 usuários que reclamaram sobre telecomunicações e que foram respondidos pelas empresas no momento solicitado. Abaixo, no gráfico 05 é possível vislumbrar esses valores em relação ao total de reclamações:

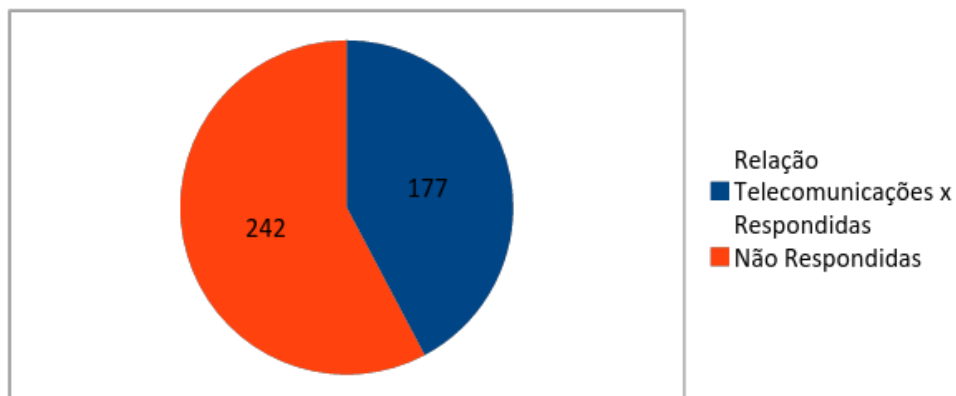


Gráfico 05: Relação entre Área de Telecomunicações e Respondidas

Na quinta regra de Associação, o algoritmo apresenta a relação existente entre a localidade do solicitante, neste caso, João Pessoa, a situação da Reclamação, que na aplicação apresentou-se como Finalizada e Avaliada e se o usuário foi respondido pela empresa quando

solicitado. O resultado apresentado pela aplicação foi que 171 usuários de João Pessoa e que Finalizaram e Avaliaram a Reclamação tiveram sua solicitação de resposta da empresa atendida. A relação proporcional desta relação pode ser expressa pelo gráfico 06 abaixo:

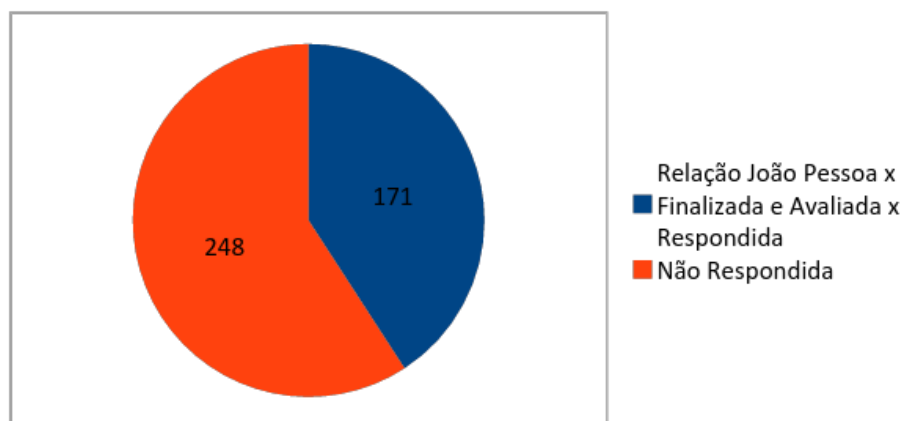


Gráfico 06: Relação João Pessoa x Finalizada e Avaliada x Respondida

Na sexta regra de associação podemos observar que existe uma relação entre as pessoas que reclamaram sobre Cobranças/Contestações e 170 dessas pessoas foram respondidas. O gráfico 07 abaixo expressa diretamente essa relação, apresentando o número respondido em relação ao grupo total do experimento:

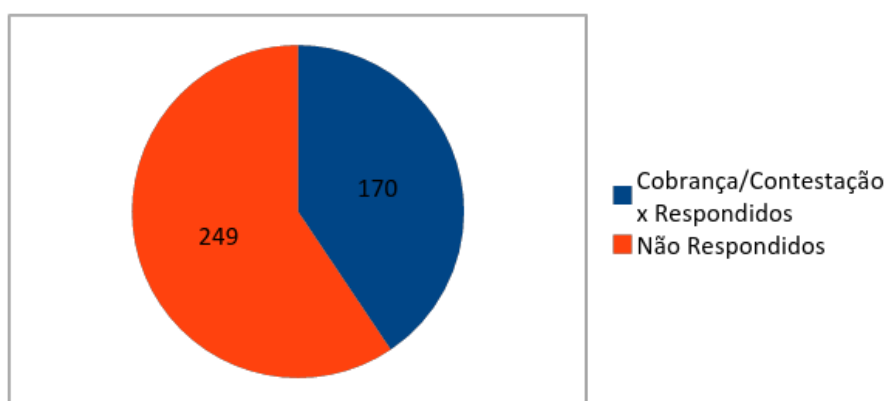


Gráfico 07: Relação Cobrança/Contestação x Respondidos

Na sétima regra de Associação gerada pelo algoritmo *Apriori* é possível observar que 169 pessoas do sexo Masculino, que procuraram a empresa reclamada e que Finalizaram e

Avaliaram a reclamação, foram respondidos pela empresa em momento oportuno. A expressão gráfica desta relação pode ser observada abaixo:

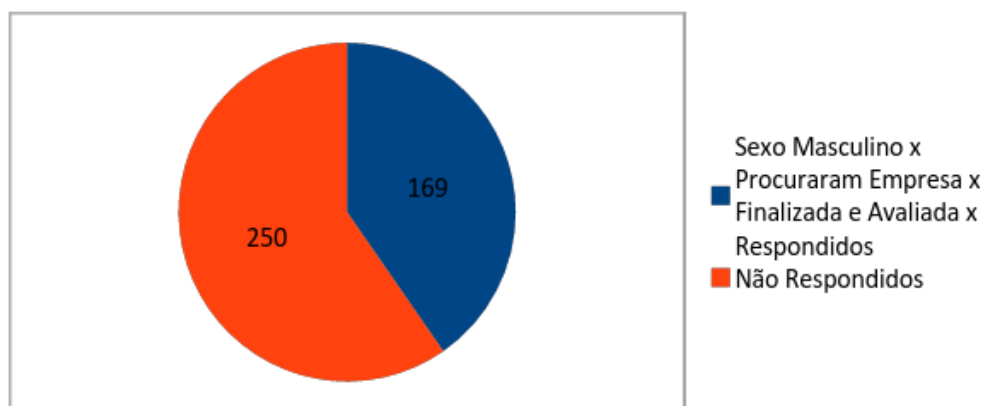


Gráfico 08: Relação Sexo Masculino x Procurou Empresa x Finaliza e Avaliada x Respondida

A oitava relação encontrada pelo WEKA demonstra que 158 pessoas que reclamaram sobre Telecomunicações e que procuraram a empresa tiveram sua reclamação respondida pela reclamada em momento oportuno. A expressão relacional pode ser observada no gráfico abaixo:

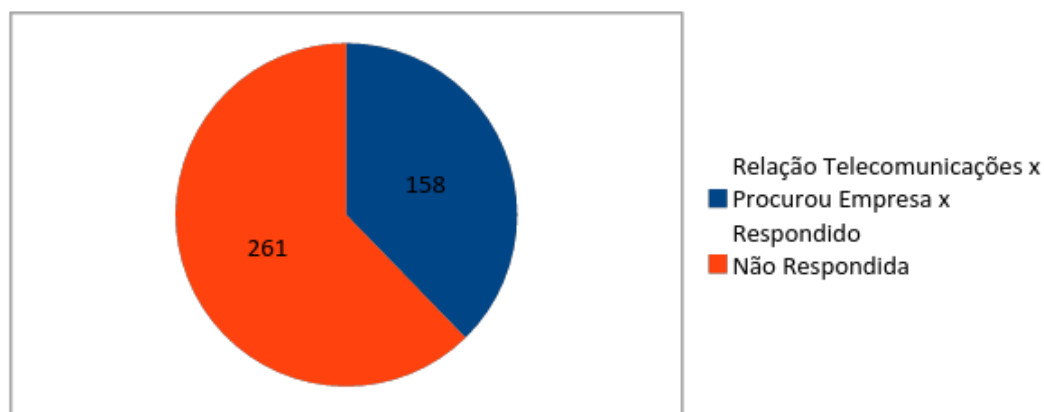


Gráfico 09: Relação Telecomunicações x Procurou Empresa x Respondido

A Nona regra de Associação descoberta pela aplicação informa que Pessoas que moram em João Pessoa, que procuraram a empresa e que Finalizaram e Avaliaram suas reclamações foram respondidos pela reclamada em momento oportuno. Temos, assim, um

total de 153 pessoas que obedecem a essa relação. Abaixo, o gráfico 10 demonstra a relação desse total em função do total de pessoas da aplicação:

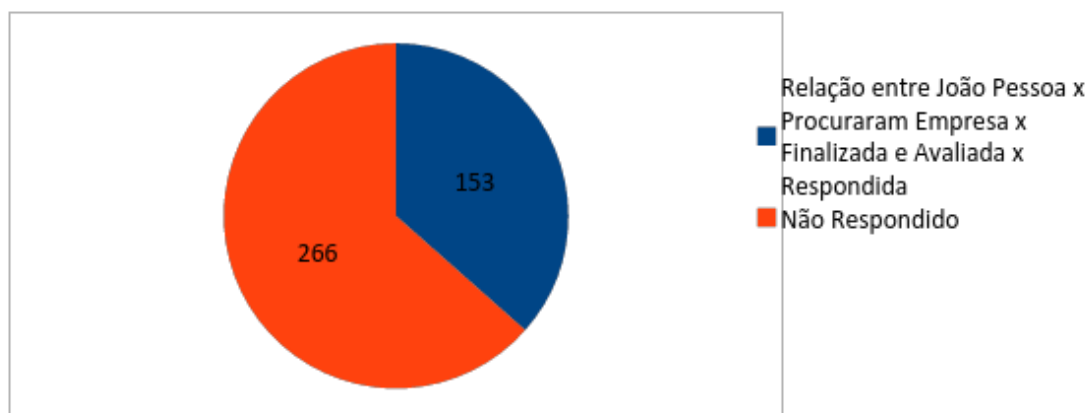


Gráfico 10: Relação entre João Pessoa x Procuraram Empresa x Finalizada e Avaliada x Respondida

Por fim, apresentamos a análise realizada pela aplicação que mostrou a décima regra de Associação informando que 162 Pessoas que tem as características de Reclamar sobre Cobranças/Contestação e que procuraram a empresa tiveram uma resposta da mesma em momento oportuno. Abaixo, o gráfico que representa esse valor em relação ao total da aplicação:

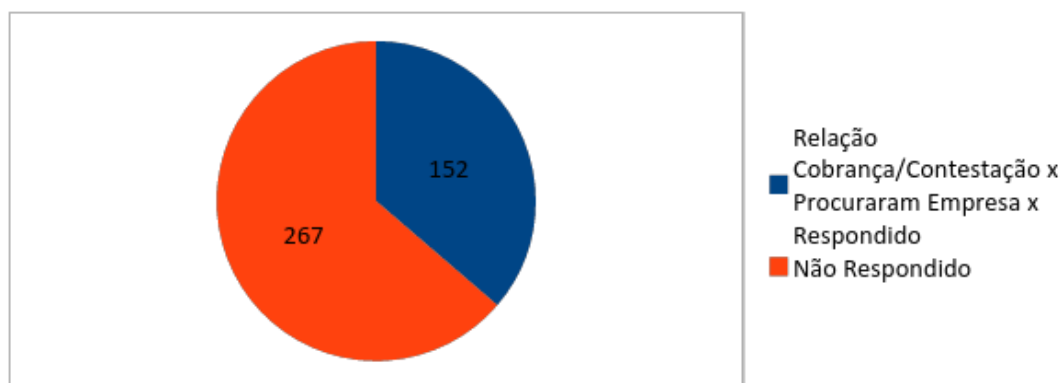


Gráfico 11: Relação Cobrança/Contestação x Procurar Empresa x Respondido

Durante todo o processo de desenvolvimento da aplicação, é possível observar que como classe base para os resultados, o objetivo foi descobrir relações que caracterizassem as respostas como respondidas. Com esse conjunto de informações, pessoas teriam condições de

realizar uma análise mais aprofundada, a fim de analisar o perfil dos usuários, a localização do maior número de respostas respondidas, a organização ou setor ou segmento de mercado que menos responde a seus usuários, se usuários do sexo masculino são mais respondidos que os do sexo feminino, e as características que o levam a ser mais respondido.

As possibilidades de análise de resultados são imensas, e as aplicações de seus resultados impactantes para organizações, gerando relatórios e gráficos detalhados, sendo assim considerada informação de impacto.

Com a apresentação dessas regras de associação e dos resultados numéricos desta aplicação, é possível observar o Impacto positivo do Business Inteligente nas organizações, e como essa ferramenta é poderosa quando bem usada em bancos de dados que, anteriormente, eram considerados sem uso ou sem importância.

6. RESULTADOS

Abaixo, é possível realizar a visualização dos resultados correlacionados com os objetivos e questões de pesquisa deste trabalho.

	Objetivos	Resolução
1	Realizar Análise do Banco e Determinar a técnica de Data Mining mais adequada, utilizando como ferramenta o Software Weka	Foi possível visualizar a seleção do banco de dados do Consumidor.gov.br, justificando o por que de sua utilização. Também foi contextualizada a técnica de Associação utilizando o algoritmo <i>Apriori</i> , demonstrando como o mesmo funciona e apresentando os fundamentos para este trabalho.
2	Realizar o processo de Mineração de Dados, seguindo as etapas prescritas e organizadas no KDD, descrevendo os resultados com a aplicação.	Durante a realização deste trabalho, pode ser observado que todas as etapas que compreendem desde a seleção até a análise da informação extraída foi seguida e apresentada, tendo como foco o processo de transformação dos dados em informação.
3	Apresentar os Resultados da Aplicação da Técnica de Associação	Neste trabalho foi possível visualizar o resultado do processo de KDD, mais especificamente, do processo de Data Mining, utilizando para isso o algoritmo <i>Apriori</i> .
4	Observar a solução dos problemas com base nas regras de associação e demonstrar a eficiência do ambiente em apoiar o consumidor;	Com a apresentação dos resultados, é visível que grande parte dos solicitantes foram atendidos pelas organizações, e com essas informações, as empresas podem analisar perfil e traçar melhores estratégias de atendimento ao consumidor, vislumbrando aqueles que foram atendidos e analisando os motivos pelos quais os outros com perfil diferentes não tiveram suas solicitações resolvidas.

Quadro 01: Resultados dos Objetivos Específicos do Trabalho

	Questões	Resolução
1	Qual a real importância do BI para o processo de tomada de decisão?	Neste trabalho, foi possível apresentar a importância do <i>business intelligence</i> para

		todo o processo organizacional, bem como as ferramentas de <i>BI</i> podem impactar positivamente para todo o processo de Tomada de Decisão.
2	Os consumidores podem considerar o ambiente do consumidor.gov como um forte meio de solução de seus problemas e comunicação com as organizações reclamadas?	Tendo em vista a análise e os gráficos apresentados, bem como os perfis dos usuários, é visível que o ambiente funciona na maioria dos casos para resolução de problemas entre consumidor e empresa.
3	Podemos avaliar a eficiência de órgãos com base em seus dados públicos ?	Com este trabalho, fica comprovada que a utilização de bases de dados públicas é extremamente eficiente no processo de análise dos órgãos públicos. Especificamente, a base de dados pública foi vista como muito positiva na análise dos dados e em sua relevância para avaliação do ambiente de reclamação.

Quadro 02: Resultados das Questões de Pesquisa do Trabalho

7. TRABALHOS FUTUROS

Com a vasta quantidade de dados e a profundidade que pode ser alcançada na pesquisa, fica visível a existência de possíveis trabalhos que possam complementar ou atualizar os dados dispostos neste texto.

Sendo assim, vislumbramos como possíveis trabalhos para o futuro um aprofundamento dos dados, podemos destacar a comparação entre os dados de diferentes estados e do país, fazendo assim um panorama de como andam as soluções de reclamações no estado da Paraíba em relação a outros estados ou ao país.

Vislumbramos também como trabalho futuro a comparação dos dados dispostos neste trabalho com dados de outras entidades especializadas em solução de problemas entre empresas e consumidores, tais como PROCON, Anatel, ou possíveis processos judiciais que possam existir. Com esse comparativo, é possível tirar um panorama se a ferramenta tem eficácia superior, igual ou inferior, medindo assim sua eficiência com relação a esses outros meios de solução de problemas.

8. CONCLUSÃO

A evolução do conceito de informação fez com que as organizações vislumbassem mais do que um mero conjunto de dados armazenados em banco, mas sim um conjunto de grandes possibilidades de evolução, que gera um considerável diferencial para a organização e apresentando-se como forte meio de obtenção de conhecimento antes não observado ou conhecido.

A utilização de *Business Intelligence* nas organizações têm apresentado resultados satisfatórios, e demonstrado ser uma ferramenta de grande impacto no processo de predileção e classificação de informações. Anteceder conhecimentos através de informações e análise de dados pode ser considerada uma tendência futura nas organizações de médio e pequeno porte e é uma realidade nas organizações de grande porte.

O Data Mining mostra-se satisfatório para diversos cenários e o conceito de aprendizado de máquina extremamente vantajoso na construção de soluções e na adaptação de novas informações com base no modelo criado, apresentando números e resultados vantajosos e com índices de acerto consideráveis, provando que a técnica é realmente válida e aplicável em diversos cenários e situações.

Com o conjunto de dados correto, podemos então definir, que a aplicação de aprendizado de máquina é o futuro nas organizações, e a receita para a análise dos grandes e abandonados dados organizacionais. Bem como para a criação de informações que auxiliem essas informações a analisarem suas relações com consumidores e criarem perfil vislumbrando o desenvolvimento de políticas de solução de problemas mais eficientes e rápidas.

REFERÊNCIAS

BARBIERI, Carlos. BI – **BUSINESS INTELLIGENCE – Modelagem & Tecnologia**, Ed. Axcel Books – 2001.

BATISTA, Emerson de Oliveira. **SISTEMA DE INFORMAÇÃO: o uso consciente da tecnologia para o gerenciamento**, Ed. Saraiva, 2004.

Boente, Alfredo Nazareno Pereira. Goldschmidt. Estrela, Vânia Vieira . **Uma Metodologia de Suporte ao Processo de Descoberta de Conhecimento em Bases de Dados**. In: V Simpósio de Excelência em Gestão e Tecnologia, Resende – RJ. V SEGeT, v. 1. p. 4-5. 2008.

CARDOSO, Olinda Nogueira Paes & MACHADO, Rosa Teresa Moreira. **Gestão do conhecimento usando data mining: estudo de caso na Universidade Federal de Lavras**. Revista de Administração Pública do Rio de Janeiro, v.42, n.3, Mai./Jun. 2008.

Dilly, R. **Data Mining: an introduction**. Belfast: Parallel Computer Centre, Queens University. 1999.

Drucker, Peter F. **O Advento da Nova Organização**. Série Harvard Business Review. Rio de Janeiro: Ed. Campus Ltda. 2001.

Fayyad, Usama. PIATETSKY-SHAPIRO, Gregory e SMYTH, Padhraic. **From Data Mining to Knowledge discovery**. American Association for Artificial Intelligence. 1996.

Goldschmidt, R E. Passos, E. **Data mining: Um guia prático**. Rio de Janeiro: Campus. 2005.

Goode, W. J.; Hatt, P. K. **Métodos em pesquisa social**. 3. ed. São Paulo: Cia. Editora Nacional. 1969

Lacerda, M. P.; Souza, R. C. F. **Aplicação da Mineração de Dados em Sistema de Avaliação de professor e aluno**. Monografia(conclusão de curso). Universidade Federal do Pará. Belém. 2004

NERY JUNIOR, Nelson. **Código Brasileiro de Defesa do Consumidor comentado pelos autores do Anteprojeto**. 4. ed. Rio de Janeiro: Forense Universitária, 1995.

NUNES, Luis Antonio Rizzatto. **Curso de Direito do Consumidor**. 4. ed. São Paulo: Saraiva, 2009. p. 2.

PETRINI, M.; POZZEBON, M.; FREITAS, M. T. **Inteligência De Negócios No Brasil**, Ed.HSM Management, 2005.

Rezende, S. O. **Sistemas Inteligentes: Fundamentos e Aplicações**. Barueri, SP, Brasil. Editora Manole. 2003.

ROCHA, Sílvio Luís Ferreira da. **Responsabilidade civil do fornecedor pelo fato do produto no direito brasileiro**. São Paulo: Revista dos Tribunais, 1992.

SEZÕES, Carlos; OLIVEIRA, José; e BAPTISTA, Miguel. **Business Intelligence**. Porto, Portugal. SPI – Sociedade Portuguesa de Inovação: 2006.

YIN, R. K. **Case study research: design and methods**. USA: Sage. 1989